Unsupervised Machine Learning framework for sensor placement optimization: analyzing methane leaks

Shirui Wang[*] University of Houston

Sara Malvar[*†] Microsoft

Leonardo O. Nunes Microsoft

Kim D. Whitehall Microsoft

Yagna D. Oruganti Microsoft

Yazeed Alaudah Microsoft

Anirudh Badam Microsoft

Abstract

Methane is one of the most potent greenhouse gases, with the global oil and gas industry being the second largest source of anthropogenic methane emissions, accounting for about 63% of the whole energy sector. This underscores the importance of detecting and remediating methane leaks for the entire oil and gas value chain. Methane sensor networks are a promising technology to detect methane leaks in a timely manner. While they provide near-real-time monitoring of an area of interest, the density of the network can be cost prohibitive, and the identification of the source of the leak is not apparent, especially where there could be more than one source. To address these issues, we developed a machine learning framework that leverages various data sources including oil and gas facilities data, historical methane leak rate distribution and meteorological data, to optimize sensor placement. The determination of sensor locations follows the objective to maximize the detection of possible methane leaks with a limited sensor budget.

1 Introduction

Methane (CH_4), the primary component of natural gas, is a potent greenhouse gas with a Global Warming Potential (GWP) of 84-87 over a 20-year timescale (1). The Intergovernmental Panel on Climate Change (IPCC) recommends, with high confidence, that reduction of anthropogenic methane emissions is an efficient way to limit global temperature rise to 1.5°C above pre-industrial levels by 2030 (2).

The global oil and gas industry is one of the primary sources of anthropogenic methane emissions, with significant leaks occurring across the entire oil and gas value chain, from production and processing to transmission, storage, and distribution (1). Capacity limitations in gathering, processing and transportation infrastructure can lead to venting/flaring of excess CH_4. The International Energy Agency (IEA) estimates that it is possible to avoid around 70% of today's methane emissions from global oil and gas operations with existing technologies (3). These statistics drive home the importance of leveraging various methane detection technologies along with machine learning to address this critical issue.

To measure CH_4 emissions, several methane-sensing instruments are used, with different sensitivities, spatiotemporal resolutions, collection techniques and cost, and frequency of collection. Various approaches have been proposed to detect methane emissions and leaks using different emerging

[*]These authors contributted equally to the paper

[†]Corresponding author: saramalvar@microsoft.com

Tackling Climate Change with Machine Learning workshop at NeurIPS 2021

methane-sensing technologies such as mobile methane detection and quantification of leaks and monitoring with satellite (4; 5), monthly and interannual aerial surveys to monitor CH_4 persistence (6; 7), FLIR cameras based on optical gas imaging technologies for detecting leakages with high confidence (8), IoT sensor grids for measuring real-time methane concentrations (9) and stationary sensors to quantify sources of emissions and for monitoring large areas of interest (10; 11). One of the technologies used to help with methane leak detection and remediation efforts (LDAR) is ground sensor networks (12). While sensors can provide realtime/near realtime measurements of methane concentration, the key challenge in employing them over a large area of interest - say, the size of the Permian Basin in Texas, USA - would be the prohibitive cost of dense sensor placement. While ground sensors can provide realtime/near-realtime measurements of methane concentration, challenges arise in deploying them in a stationary configuration for maximum detection of leaks within an area of interest that may contain multiple sources and establishing attribution to a given source based on sensor readings. This underscores the importance of a robust sensor placement optimization method to support LDAR efforts. To tackle this task, Klise et al. (11) proposed an optimization strategy leveraging atmospheric dispersion and transport models and developed an open-source Python package named Chama. The maximum-coverage problem is formulated based on simulated methane dispersion maps and solved by mixed-integer linear programming formulations, considering the ingested sensor data, such as sensitivities, number of sensors and budget. This strategy starts from initialized sensor locations and finds the optimal subset as the result, but it is highly dependent on the initialization of the sensor locations, that can only be placed on a specific subset of initial positions. In our framework, the maximum-coverage problem is solved using a machine learning (ML) unsupervised approach.

In this paper, we propose a framework to incorporate various oil and gas assets, leak/methane emission history, meteorological data, topographical data, and pollutant dispersion and transport into machine learning methods to suggest a sparse sensor placement in a stationary network for an area of interest. The objective is to maximize the detection of all emissions sources and facilitate leak source attribution, thereby helping mitigate and remediate methane emissions in an accurate and timely manner.

2 Proposed sensor placement optimization method for methane leak detection

The major innovation of the proposed method is that it provides an unsupervised learning solution by leveraging the methane geospatial features generated from dispersion and transport models. It allows different kinds of heterogeneous data sources as the inputs, varying from asset location, asset properties to weather information, and underlying surface imagery. We followed the problem formulation proposed by Klise et al. (11) and construct the framework based on Figure 1, composed of three core processes. The first two steps demonstrate similar modeling processes as outlined in Klise et al. (11) and prepare the methane geospatial features for the optimization. Instead of adopting mixed integer linear programming, we proposed a new ML based approach for solving the optimization problem.

The proposed three core steps are:

1. Data ingestion: This step relates to acquiring and ingesting the data required for the solution. Oil and gas facilities locations data, including wells, natural gas pipelines and processing plants, available in the public domain, are ingested for the area of interest (13; 14). One can also infer oil and gas facilities locations, such as tank batteries, using high spatial resolution satellite imagery through image segmentation techniques, given sufficient labeled data.

   Historical emissions distribution data related to source leak rates, which is a function of the type of oil and gas facility in question and is based on data from an extensive airborne campaign across the Permian Basin from September to November of 2019 (15), is also ingested at this time. Further, as these sources can have varying heights and land topography may impact the movement of the methane concentration in time, a digital elevation map is also ingested in this step. Highly spatiotemporally resolved meteorological data at the surface as well as within the atmosphere for the area of interest is also required. Ideally, these data should be temporally resolved at a minimum of diurnally and be spatially representative of the microclimate within the area of interest. These data may be point sourced or gridded.

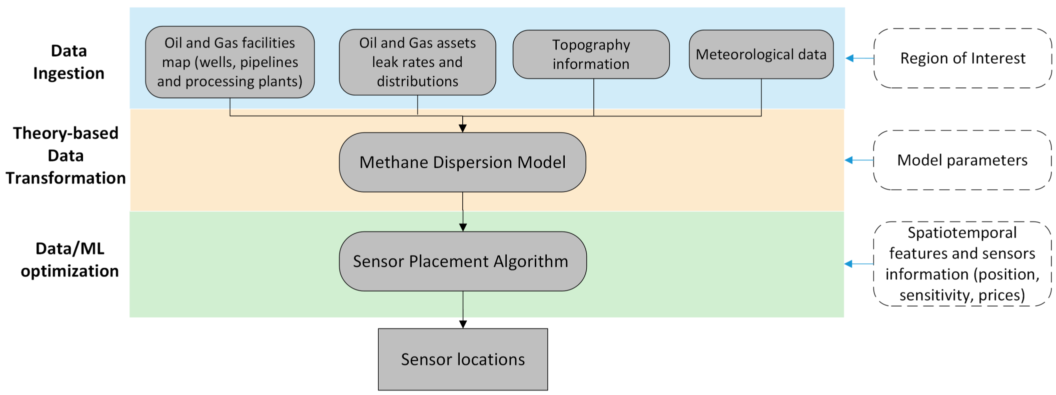

Figure 1: The proposed framework is composed of three steps: data ingestion, data transformation and machine learning optimization. The data ingestion step ingests the data required for the modeling and optimization components of the solution. In the next step, the results of several leakage scenarios within an area of interest are simulated. In the final step, machine learning is applied to determine the optimum location of the sensor, constraining on sensor information such as budget and sensitivity, or on methane plume feature properties, or a combination of both.

Data related to the sensors to be used for the proposed network within the area of interest are also ingested at this step. These data may be related to sensor cost, detection sensitivity as outlined in Klise et. al (11).

2. Methane geospatial feature generation: This is the theory-driven aspect of the framework that models the movement of CH_4 in the atmosphere over the area of interest within the planetary boundary layer (up to 1.5km into the atmosphere) using atmospheric-based dispersion models. These models may range from Large Eddy Simulation and the Gaussian plume models, to atmospheric dispersion models and numerical weather prediction models with radiative chemistry and transport schemes. The choice of model to use should be driven by the complexity of the microclimate within the area of interest, the model accuracy required, and the resources available to address the computational requirements of the model.
3. Integrate machine learning to identify the locations for sensor placement: Once the features have been determined using the physics-based atmospheric dispersion models and sensor placement optimized based on the sensor requirements for the network, the idea is to find spatial redundancy for sensor locations that capture the variance of CH_4 geospatial features over days. The coverage formulation is used which selects a set of sensors that maximizes an objective. Two objectives can be used: (I) scenario coverage, which maximizes the number of scenarios detected, and (II) scenario-time coverage, which maximizes the average amount of time each scenario is detected. First, the optimization algorithm selects grid points with maximum scenario coverage. Then, the points are spatially clustered using DBSCAN and we get the centroid of the largest cluster. The covered scenario is then removed from the simulation and the process repeated till the assigned number of sensors.

In a case study, preliminary results showed that the methodology proposed surpasses the baseline model, in which sensors are placed along the downwind direction of each source. The proposed cluster-greedy method detected 87.9% of the CH_4 leaking sources, while the baseline model detected 82.8%, a performance improvement of 5.8%. Using only one sensor for every three possible emission sources, the improvement in performance is even more pronounced. In this scenario, the methodology proposed in this paper detected 6.79% more leaks than the baseline model.

3 Summary

In this work, we proposed a framework for sensor placement optimization for an area of interest, using the maximum coverage formulation given a sensor budget, by leveraging physics-based atmospheric dispersion model outputs, and incorporating various data sources such as oil and gas facility maps, weather data, and distribution of historical methane leak rates.

References

[1] H. Conick, A. Revi, M. Babiker, P. Bertoldi, M. Buckeridge, A. Cartwright, W. Dong, J. Ford, S. Fuss, J.-C. Hourcase, D. Ley, R. Mechler, P. Newman, A. Revokatova, S. Schultz, L. Steg, and T. Sugiyama, "Strengthening and implementing the global response.," Tech. Rep. In Press., In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty, [Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P.R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M.I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, and T. Waterfield (eds.)], 2018.

[2] J. Rogelj, D. Shindell, K. Jiang, S. Fifita, V. Ginzburg, C. Handa, H. Kheshgi, S. Kobayashi, E. Kriegler, L. Mundaca, R. Séférian, and M. Vilariño, "Mitigation pathways compatible with 1.5°C in the context of sustainable development," 2018.

[3] IEA, Paris, "Driving down methane leaks from the oil and gas industry." `https://www.iea.org/reports/driving-down-methane-leaks-from-the-oil-and-gas-industry`, 2021. Accessed: 2021-09-13.

[4] T. G. van Kessel, L. J. Klein, M. Ramachandran, E. J. Zhang, and H. Hamann, "Satellite guided mobile wireless methane detection for oil and gas operations," in *2020 IEEE 6th World Forum on Internet of Things (WF-IoT)*, pp. 1–4, 2020.

[5] J. Mckeever, D. Jervis, and M. Strupler, "Microsatellites spot mystery methane leaks," *IEEE Spectrum*, vol. 57, no. 11, pp. 38–43, 2020.

[6] J. Englander, A. Brandt, S. Conley, and D. Lyon, "Aerial interyear comparison and quantification of methane emissions persistence in the Bakken formation of North Dakota, USA.," *Environmental Science & Technology*, vol. 52, no. 15, pp. 8947–8953, 2018.

[7] D. Lyon, R. Alvarez, D. Zavala-Araiza, A. Brandt, R. Jackson, and S. Hamburg, "Surveys of elevated hydrocarbon emissions from oil and gas production sites.," *Environmental Science & Technology*, vol. 50, no. 9, pp. 4877–86, 2016.

[8] A. P. Ravikumar, J. Wang, M. McGuire, C. S. Bell, D. Zimmerle, and A. R. Brandt, "Good versus good enough? empirical tests of methane leak detection sensitivity of a commercial infrared camera," *Environmental Science & Technology*, vol. 52, no. 4, pp. 2368–2374, 2018.

[9] L. Klein, R. Muralidhar, F. Marianno, J. Chang, S. Lu, and H. Hamann, "Geospatial internet of things: Framework for fugitive methane gas leaks monitoring," *International Conference on GIScience Short Paper Proceedings*, vol. 1, 01 2016.

[10] R. Heltzel, *On the Improvement of the Indirect Quantification of Methane Emissions: A Stationary Single Sensor Approach*. PhD thesis, 8329, 2021.

[11] K. Klise, B. Nicholson, C. Laird, A. Ravikumar, and A. Brandt, "Sensor placement optimization software applied to site-scale methane-emissions monitoring," *Journal of Environmental Engineering*, vol. 146, p. 04020054, 2020.

[12] L. Klein, "Geospatial internet of things: Framework for fugitive methane gas leaks monitoring.," *International Conference on GIScience Short Paper Proceedings*, 2016.

[13] Railroad Commision of Texas, "Data sets available." `https://www.rrc.texas.gov/resource-center/research/data-sets-available-for-download/`, 2021. Accessed: 2021-09-13.

[14] U.S. Energy Information Administration, "Layer information for interactive state maps." `https://www.eia.gov/maps/layer_info-m.php`, 2020. Accessed: 2021-09-13.

[15] D. H. Cusworth, R. M. Duren, A. K. Thorpe, W. Olson-Duvall, J. Heckler, J. W. Chapman, M. L. Eastwood, M. C. Helmlinger, R. O. Green, G. P. Asner, P. E. Dennison, and C. E. Miller, "Intermittency of large methane emitters in the Permian basin," *Environmental Science & Technology Letters*, vol. 8, no. 7, pp. 567–573, 2021.